

Máster en Programación avanzada en Python para Big Data, Hacking y Machine Learning

Programación Python para Machine Learning

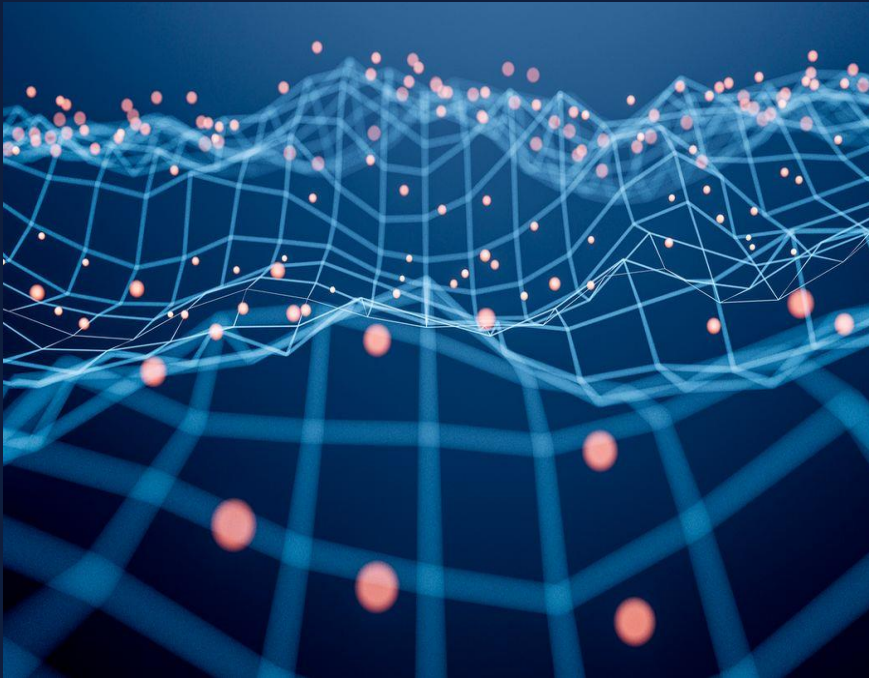
LECCIÓN 06

Lección 6: Regresión Lineal, Regresión Logística y kNN.

ÍNDICE

- ✓ Introducción
- ✓ Objetivos
- ✓ Regresión Lineal: principios e implementación
- ✓ Regresión Logística: principios e implementación
- ✓ K-Vecinos más cercanos: principios e implementación
- ✓ Consideraciones a tener en cuenta.
- ✓ Conclusiones

INTRODUCCIÓN



- ✓ Técnicas básicas y simples.
- ✓ No están exentas de potencia.
- ✓ Problemas que pueden resolver.

OBJETIVOS

Al finalizar esta lección serás capaz de:

- 1 Conocer los principios teóricos de la Regresión Lineal y Regresión Logística.
- 2 Implementar en Python una Regresión Lineal y Logística para problemas de regresión y clasificación, respectivamente.
- 3 Conocer las bases conceptuales de los modelos k-Vecinos más cercanos.
- 4 Implementar en Python modelos de k-Vecinos más cercanos para problemas de regresión y clasificación.

REGRESIÓN LINEAL

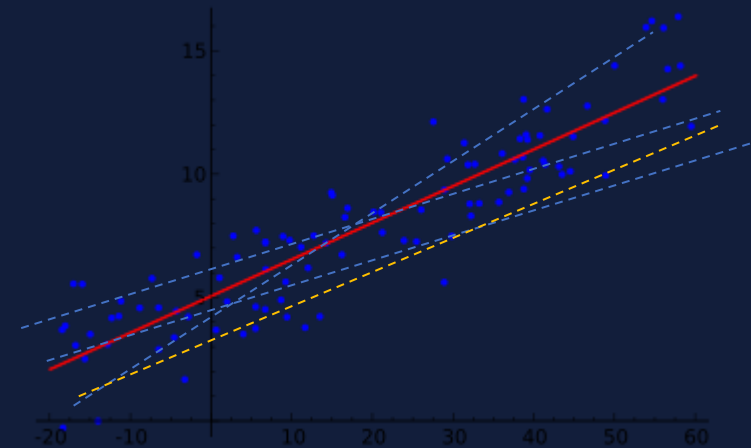
Procedimiento estadístico que permite encontrar una relación lineal entre dos variables mediante el cálculo de una recta.

Concepto de linealidad.

Ecuación de la línea recta en dos dimensiones.

$$y = wx + b$$

Objetivo: encontrar el valor óptimo para b y w a partir de una serie de puntos (x,y)

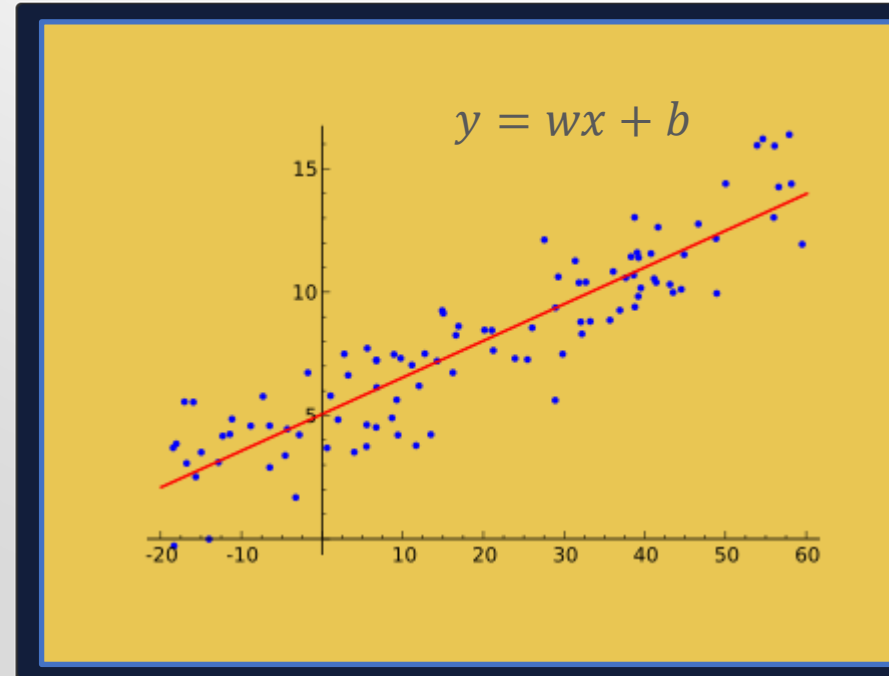


REGRESIÓN LINEAL

La regresión lineal simple no tiene porqué buscar valores mediante un cálculo analítico. Es posible estimar w y b de modo directo utilizando los datos disponibles:

$$w = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{Cov(x, y)}{Var(x)}$$

$$b = \bar{y} - w\bar{x}$$



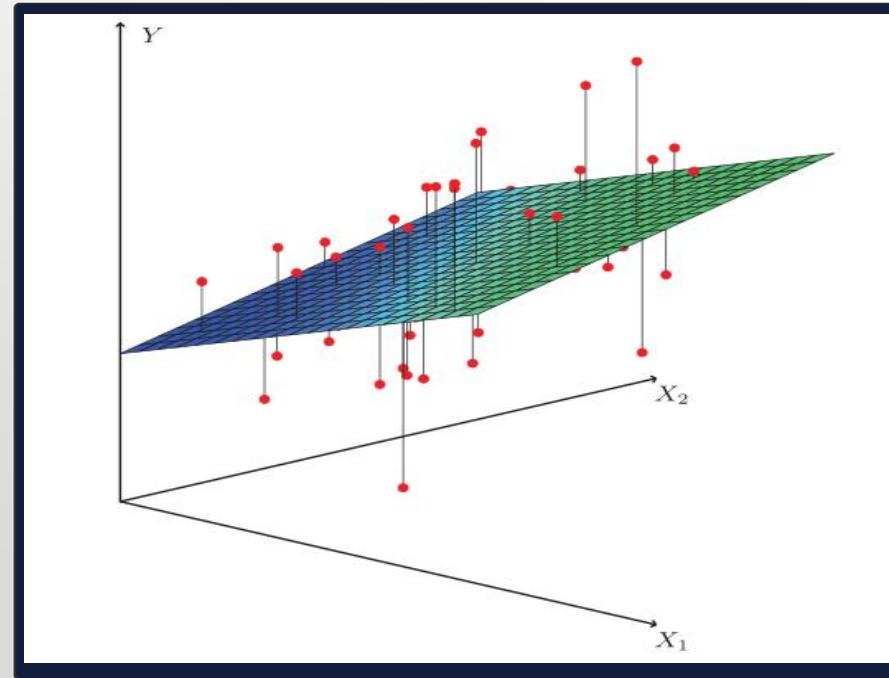
REGRESIÓN LINEAL

La regresión lineal múltiple

No tiene porqué buscar valores mediante un cálculo analítico. Es posible estimar w y b de modo directo utilizando los datos disponibles:

$$y = w_1x_1 + w_2x_2 + \dots + w_nx_n + w_0$$

Resolución: Mínimos cuadrados ordinales.



REGRESIÓN LINEAL

Modelo paramétrico:

- ✓ Independencia entre las variables entrada:
Las variables de entrada deben ser independientes, no debe de haber **colinealidad** entre ellas.
- ✓ Distribución normal de la variable de salida.
- ✓ No autocorrelación: Los valores de cada instancia para cada variable son independientes del valor de las variables de las demás instancias. Especialmente importante con datos temporales.

La *colinealidad* es un término para referirse a la situación en la que una variable de entrada está linealmente relacionada con una o varias de las otras variables de entrada del modelo.

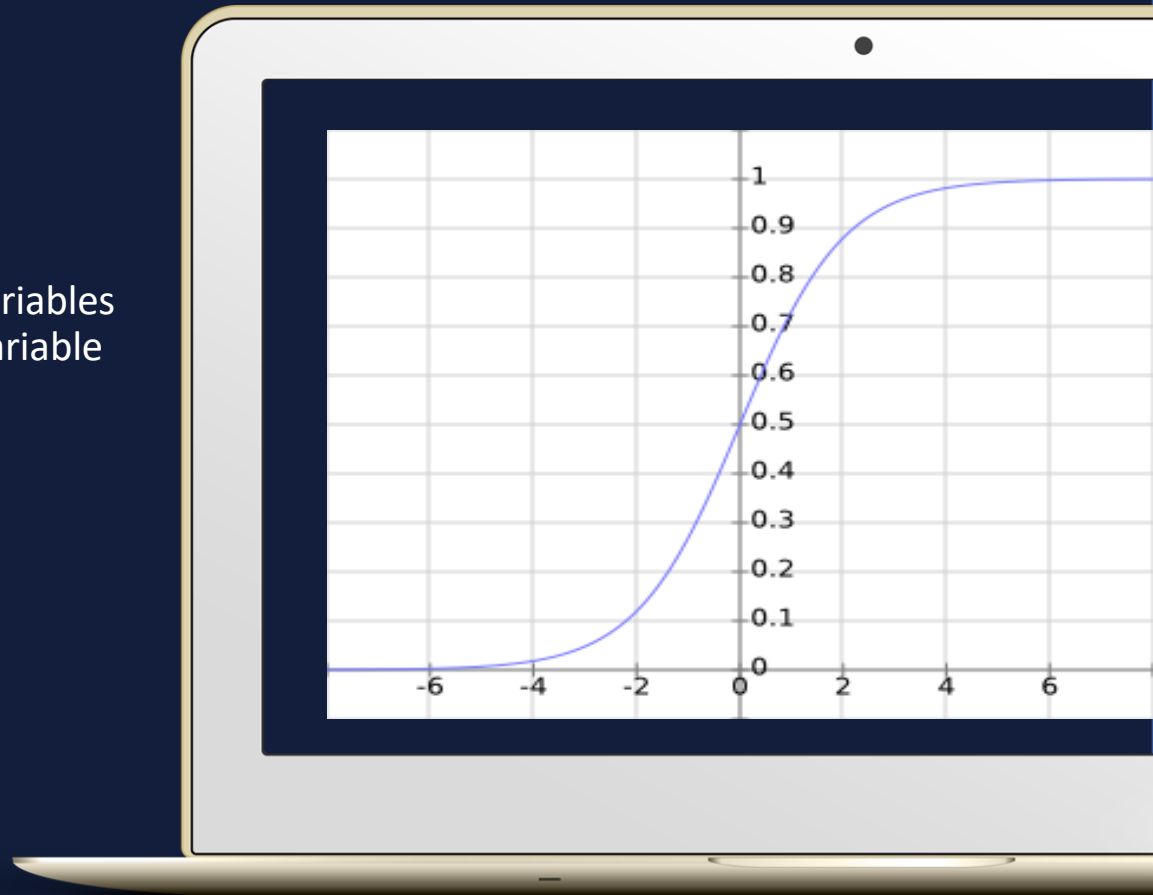
REGRESIÓN LOGÍSTICA

Modelo supervisado lineal.

Problemas de clasificación.

Trata de modelar mediante una o más variables independientes la probabilidad de una variable dependiente binaria.

$$\text{Sigmoide}(x) = \frac{1}{1 + e^{-x}}$$

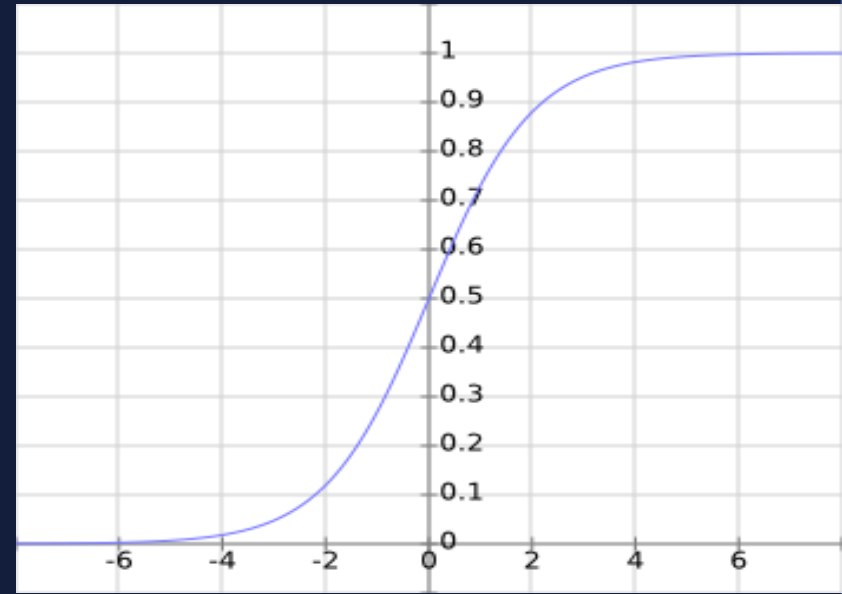


REGRESIÓN LOGÍSTICA

Regresión logística simple o múltiple.

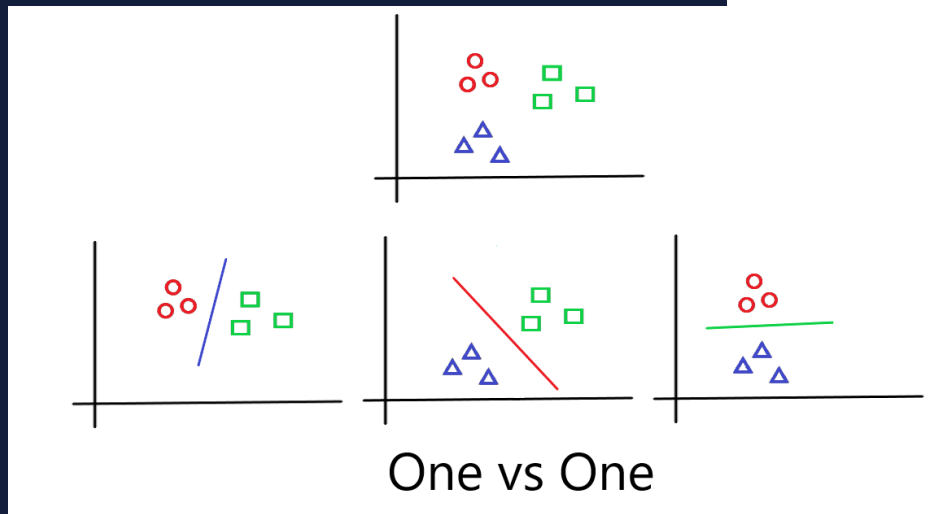
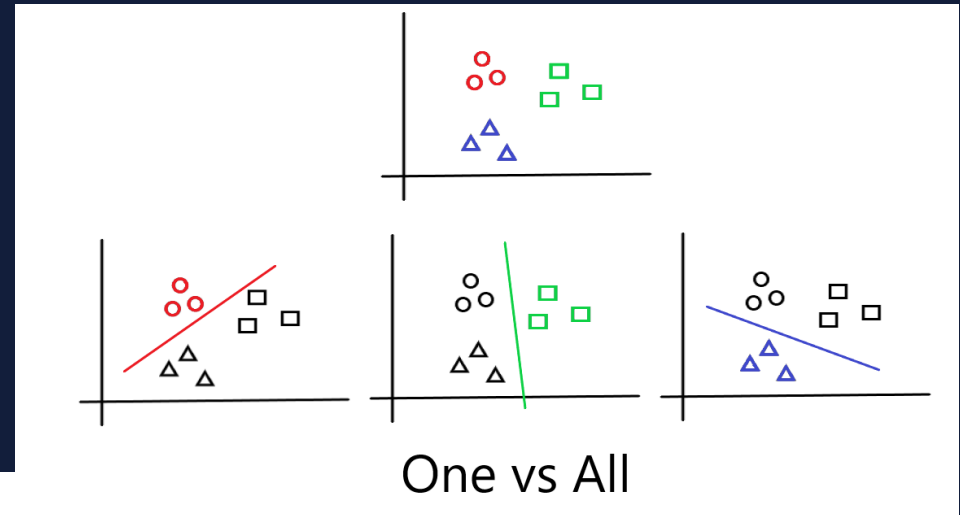
$$P(y = 1 | X = x) = \frac{1}{1 + e^{-(w_1x_1 + w_2x_2 \dots + w_nx_n + w_0)}}$$

Técnica paramétrica.

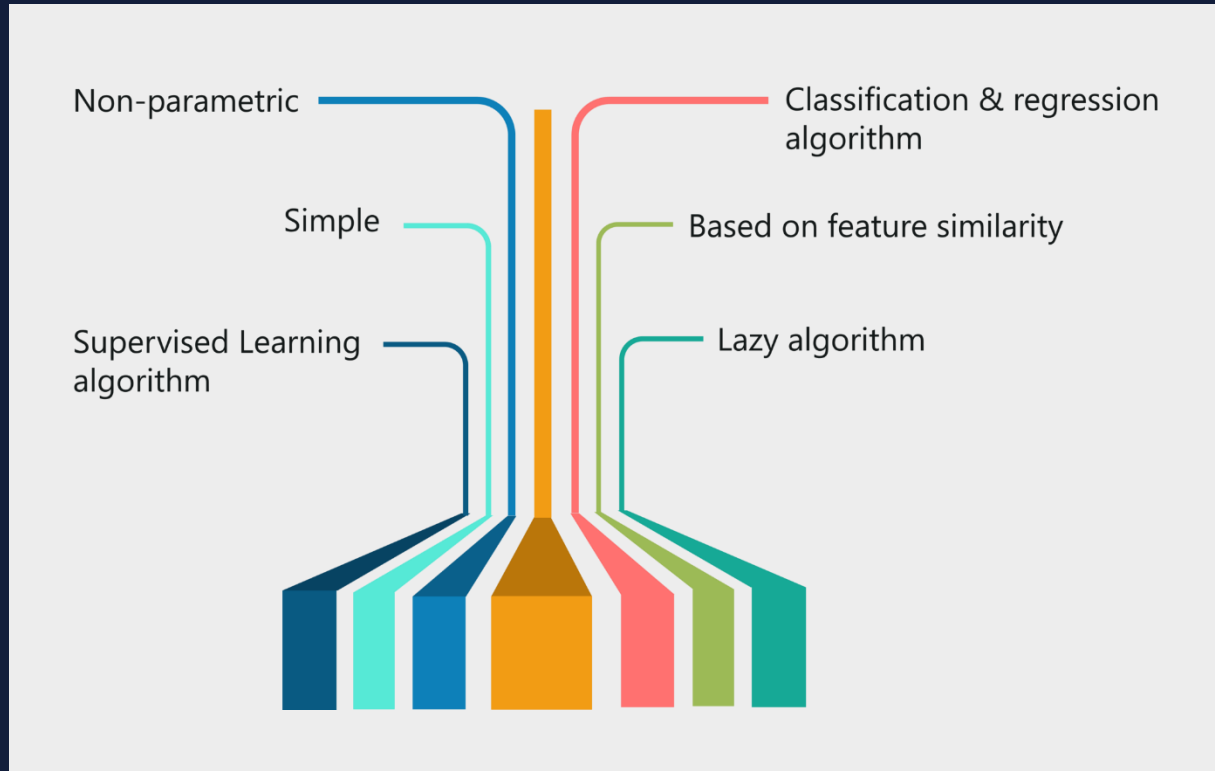


REGRESIÓN MULTICLASE

- ✓ Estrategia OVA (OVR)
- ✓ Estrategia OVO



K-Nearest Neighbors

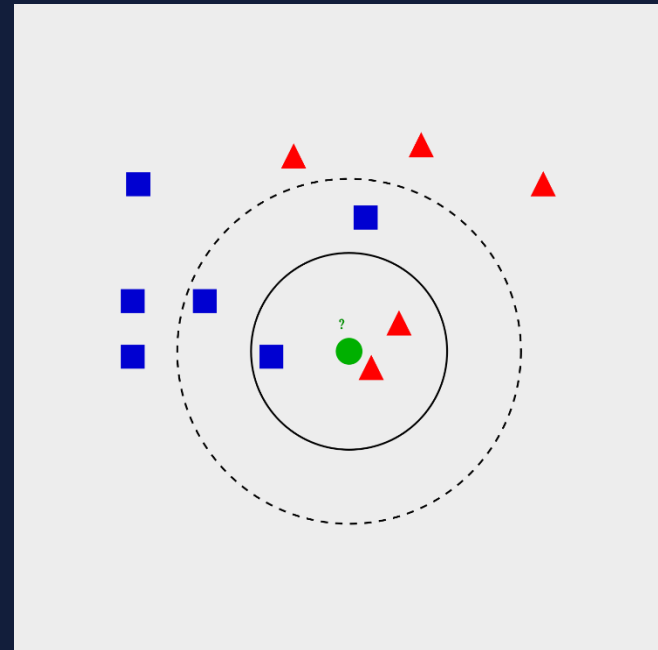


- ✓ Técnica supervisada no lineal.
- ✓ Simple pero muy potente.
- ✓ Perezoso: no entrenamiento

K-Nearest Neighbors

Fase de predicción:

1. Distancia a todos los puntos.
2. Seleccionar los k patrones más cercanos.
3. Combinar los patrones seleccionados.



MUCHAS GRACIAS POR SU ATENCIÓN



jperez@grupomainjobs.com



Javier Pérez Rodríguez
www.linkedin.com/in/perezxavi



twitter.com/eiposgrados



facebook.com/eiposgrados



instagram.com/eiposgrados